

Heart Disease Detection Using Data Mining techniques

Akhila Anikumar^{1*}, Shajan P X²

^{1,2}Sree Narayana Gurukulam College of Engineering, India

Available online at: www.ijcseonline.org

Abstract—Data mining is process to analyses number of data sets and then extracts the meaning of data. It helps to predict the patterns and future trends, allowing business in decision making. Data mining provides methods and techniques for transformation of the data into useful information for decision making. These techniques can make process fast and take less time to predict the heart disease with more accuracy. In this paper we survey different papers in which one or more algorithms of data mining used for the prediction of heart disease. By Applying data mining techniques to heart disease data which requires to be processed, we can get effective results and achieve reliable performance which will help in decision making in healthcare industry. It will help the cardiologist to diagnose the disease in less time and predict probable complications well in advance.

Keywords—Data mining, disease prediction, KNN, Decision tree, SVM

I. INTRODUCTION

Data mining is the analytical process to explore specific data from large volume of data. It is a process that finds previously unknown patterns and trends in databases. This information is further used to build predictive models. The main objective is to learn the different data mining techniques / algorithms which are used in the prediction of heart diseases using any data mining tool. Heart is the most vital part of the human body as life is dependent on efficient working of heart. A Heart disease is caused due to narrowing or blockage of coronary arteries. This is caused by the deposition of fat on the inner walls of the arteries and also due to build up cholesterol.

Heart diseases can be caused due to number of factors:

High blood pressure: when the heart pumps blood, the force of the blood pushes against the walls of the arteries causing pressure. If the pressure rises and stays high over the time it is called high blood pressure or hypertension which can harm the body in many ways i.e. increasing the risk of heart stroke or developing heart failure, kidney failure etc.

High cholesterol: cholesterol is a waxy substance found in the fatty deposits in the blood vessels. Increase in the fatty deposits (high cholesterol) does not allow sufficient blood to flow in through the arteries causing heart attacks.

Unhealthy diet: eating too much fast food increases blood pressure and cholesterol level causing the risk of heart attacks.

Smoking: it damages the lining of arteries and builds up a fatty material called atheroma which narrows the arteries causing heart attacks.

Lack of physical activity: lack of exercise increases cholesterol level in blood vessels which further increases the risk of heart attacks.

Obesity: obese people are more likely to have high blood pressure, high cholesterol level and diabetes (increase in blood sugar level) which increases the risk of heart strokes in human body. These days, data mining is gaining popularity in health care industry as this industry generates large amount of complex data about hospital resources, medicines, medical devices, patients, disease diagnosis etc. This complex data needs to be processed and analysed for knowledge extraction which will further help in decision making and is also cost effective. World health organization has estimated 17.5 million people died from cardio vascular diseases in 2012, representing 31 percent of all global deaths. Out of these, an estimated 7.4 million were due to coronary heart disease and 6.7 million were due to stroke. WHO estimated by 2030, almost 23.6 million people will die due to heart disease as written in [1]. Thus, a beneficial way to predict heart diseases in health care industry is an effective and efficient heart disease prediction system. This system will find human interpretable patterns and will determine trends in patient records to improve health care.

II. DATA MINING METHODS

There are various algorithms which can be used to predict the heart disease with accurate data.

1. MAFIA – MAFIA stands for Maximal Frequent Itemsets. Item set of Frequent is one of the fundamental data mining problems which has a goal to find the number of items which include the frequently in dataset. The major goal is to find interesting patterns from data warehouse in number

of data mining tasks like as association rules, cluster classifiers, sequence and many more. The newest method MAFIA exploits an effective algorithm which combines the ideas of old and latest algorithm to configure a realistic algorithm. The algorithm can also use for maximal mining frequent item set for searching with effective pruning algorithm.

2. K-MEANS CLUSTERING – It is an iterative and one of the best unsupervised learning algorithm to divide a given set of data in predefined set of k cluster where k is marked as input variable to solve the conventional clustering problem. The K-means algorithm is a method and frequent to use in medical area and their associated fields. K-means clustering chooses points in multidimensional space to symbolize each k cluster called centroids. A centroid is the point whose ordinates are accepted by evaluating the average of each co-ordinates of samples point that allocated the clusters. The major objective of using k-means clustering to emphasize the overall squared error function or intra-cluster deviation.

III. DATA MINING AND ITS TECHNIQUES

1. DATA MINING

It is main concerned with extracting useful information from large amount of databases. Data mining techniques and tools are used to find unknown patterns and trends from the data set. Its main objective is to automatically find the patterns in the dataset with minimal user effort and input. Data mining's main contribution is in decision making and in forecasting future trends of market. Many organisations use data mining as a tool these days for data analysis as it easily evaluates patterns and trends of market and produce effective results.

2. Data mining techniques:

2.1 Association: is one of the best known data mining technique. In association, a pattern is exposed based on a relationship of a particular item on other items in the same operation. For example, the association technique is used in heart disease prediction as it say to us the relationship of dissimilar attributes used for analysis and sort out the patient with all the risk factor which are necessary for prediction of disease.

2. CLASSIFICATION is a classic data mining technique based on machine learning. Mainly classification is used to classify every item in a set of data into one of predefined set of classes or groups. Classification technique makes use of mathematical techniques such as decision trees, linear programming, neural network and statistics.

2.3. CLUSTERING: is a data mining technique that makes significant or helpful cluster of substance that have similar feature using mechanical technique. Dissimilar from classification, clustering technique also defines the classes

and put objects in them, as in classification objects are assigned into predefined classes. For example in prediction of heart disease by using clustering obtain cluster or state that list of patients which have same risk factor. Funds this makes the split list of patients with high blood sugar and related risk factor n so on.

2.4. PREDICTION: is one of a data mining techniques that discovers relationship between independent variables and relationship among dependent and independent variables. For example, prediction analysis technique can be used in sale to predict profit for the future if consider sale is an independent variable, profit could be a dependent variable. Then based on the historical sale and profit data and can draw a fixed regression curve that is used for profit prediction.

2.5. SEQUENTIAL PATTERNS: It is a data mining technique that discovers similar patterns or regular events in transaction data over a business period.

IV. LITERATURE REVIEW

During the last decade, numerous works have been done related to heart disease prediction system using different data mining algorithms by different authors. They tried to achieve efficient methods and accuracy in finding out diseases related to heart by their work including datasets and different algorithms along with the experimental results and future work that can be done on the system to achieve more efficient results. In this paper aims at analyzing different data mining techniques that has been introduced in recent years for heart disease prediction system by different authors.

M. A. Nishara Banu *et al.* [2] used C4.5 algorithm, MAFIA and K-means clustering in the year 2014 using 13 attributes in the dataset achieving 89% accuracy. M Kumari diagnoses the heart disease by applying a data mining classifier that is Decision Tree. The research scholar analyzes the presentation of this algorithm on various factors that is accuracy rate, sensitivity and error rate. He concludes that the accuracy of Decision Tree is 79 % [16]. Dinarević *et al.* by using classification technique that constructed on supervised machine learning procedures. The author use the decision tree that has error rate 0.2775 and having accuracy of 79.05% [17]. Aqueel Ahmed *et al.* [3] show the classification techniques in data mining and show the performance of classification among them. In this classification accuracy among these data mining has discussed. In this decision tree and SVM perform classification more accurately than the other methods and was able to achieve 91% accuracy. Nidhi Bhatla *et al.* [6] projected the study of different data mining techniques that can be employed in automated heart disease prediction systems. The analysis shows that neural network with 15

attributes has shown the highest accuracy. On the other hand, Decision tree has also performed well with 99.62% accuracy by using 15 attributes.

Table-1 : Accuracy Level for each Datamining Algorithms

Data Mining Algorithm	Accuracy
C4.5 algorithm, MAFIA and K-means clustering	89%
Decision Tree	79%
Decision tree and SVM perform classification	91%
Decision tree with 15 attributes	99.62%

V. PROBLEM FORMULATION

Presently various algorithm is available for clustering the pre-processed data, in the existing work they used K-means clustering and MAFIA algorithm for Heart disease prediction system and achieved the accuracy of 89%, Decision Tree and SVM perform achieved an accuracy of 91%, Decision Trees with 15 attributes showed an accuracy of 99.62% and we can see that there is vast scope of improvement, in our proposed system we will implement SVM Classifier, Genetic Algorithm optimization over the data and set the number of attributes so as to achieve the accuracy more than the present algorithm.

Objective

1. To study the various algorithm of clustering and classifying data
2. To implement the K-mean, SVM and MAFIA algorithm
3. To implement the improved algorithm of clustering
4. Performance analysis of improved system

VI. CONCLUSION

In medical field, Data Mining provides various techniques and have been widely used in clinical decision support systems that are useful for predicting and diagnosis of various diseases. These data mining techniques used in heart diseases takes less time and make process fast for the prediction system to predict heart diseases with good accuracy in order to improve their health. In our proposed system that we are designing, will implement the improved algorithm of clustering which will allow us to achieve more accuracy than the present algorithm.

REFERENCES

- [1] Ms. Chaitrali S. Dangare, Dr. Mrs. Sulabha S. Apte, "A data mining approach for prediction of heart disease using neural networks, international journal of computer engineering and technology", 2012
- [2] M.A.Nishara Banu and B.Gomathy, "Disease Forecasting System Using Data Mining Methods", 2014
- [3] Aqueel Ahmed, Shaikh Abdul Hannan, "Data Mining Techniques to Find Out Heart Diseases", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-1, Issue-4, September 2012.
- [4] Ms. Ishtake S.H, Prof. Sanap S.A., "Intelligent Heart Disease Prediction System Using Data Mining Techniques", International J. of Healthcare & Biomedical Research, 2013
- [5] Chitra R and Seenivasagam V, "Review of Heart Disease Prediction System Using Data Mining And Hybrid Intelligent Techniques", Issn: 2229-6956(Online) Ictact Journal On Soft Computing, July 2013, Volume: 03, Issue: 04, 2013
- [6] Nidhi Bhatla and Kiran Jyoti, "An Analysis of Heart Disease Prediction using Different Data Mining Techniques", International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, Vol. 1 Issue 8, October – 2012
- [7] Shadab Adam Pattekari and Asma Parveen, prediction system for heart disease using naïve bayes, International Journal of Advanced Computer and Mathematical Sciences, 2012
- [8] Venkatadri.M, Dr. Lokanatha C. Reddy a review on data mining from past to the future. International Journal of Computer Applications, 2011.
- [9] Abhishek taneja, Heart Disease Prediction System Using Data Mining Techniques, Oriental Scientific Publishing Co., India, 2013.
- [10] Rashedur M. Rahman, Farhana Afroz, Comparison of Various Classification Techniques Using Different Data Mining Tools for Diabetes Diagnosis, Journal of Software Engineering and Applications, 2013
- [11] Nidhi Bhatla Kiran Jyoti, An Analysis of Heart Disease Prediction using Different Data Mining Techniques, International Journal of Engineering Research & Technology (IJERT), 2012.
- [12] Humar Kahramanli, Novruz Allahverdi, Design of a hybrid system for the diabetes and heart diseases, Elsevier, 2008.
- [13] Marcel A.J. van Gerven, Predicting carcinoid heart disease with the noisy-threshold classifier, Elsevier, 2007.
- [14] Mohammad Taha Khan, Dr. Shamimul Qamar and Laurent F. Massin, A Prototype of Cancer/Heart Disease Prediction Model Using Data Mining, International Journal of Applied Engineering Research, 2012
- [15] M.Akhil jabbar, Dr.Priti Chandra, Dr.B.L Deekshatulu, Heart Disease Prediction System using Associative Classification and Genetic Algorithm, International Conference on Emerging Trends in Electrical, Electronics and Communication Technologies, 2012
- [16] National High Blood Pressure Education Program Working Group on High Blood Pressure in Children and Adolescents (2004). The fourth report on the diagnosis, evaluation, and treatment of high blood pressure in children and adolescents; Paediatrics 114:555-76
- [17] Dinarević S, Mesihović H, Simeunović S, Zulić I (1994) Dyslipoproteinaemia in Children with Heart Disease. Intercontinental Cardiol 3:126-9.